

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 25 日現在

機関番号：13801

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11180

研究課題名（和文）相違写像の超距離及びその一般化による近似理論とアルゴリズムの開発

研究課題名（英文）Approximation of dissimilarity mappings by ultrametrics and their generalization: Theory and algorithms

研究代表者

安藤 和敏（Ando, Kazutoshi）

静岡大学・工学部・教授

研究者番号：00312819

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：相違写像が与えられたときに、この行列を超えない L_p ノルムに関して最も近い超距離を求める問題（ L_p -最小増加超距離問題）は系統学における重要な課題である。この問題は p が無限大であるときは効率的に解くことができるが、 p が有限のときはNP困難である。本研究では $p=1,2$ に対する L_p -最小増加超距離問題に対する局所探索アルゴリズムを開発した。

閉路完全距離は超距離の一般化である。本研究では、閉路完全距離に対して関連する集合族による特徴付けを与えた。さらに閉路完全距離の一般化である k -連結完全距離を導入し、任意の相違写像を k -連結完全距離によって L_p ノルムに関して近似するための効率的なアルゴリズムを与えた。

研究成果の学術的意義や社会的意義

p が有限の場合の L_p -最小増加超距離問題に対して分枝限定法や近似アルゴリズムが提案されているが、これらは計算時間の観点から解の品質に関して満足できるものではない。本研究で開発したこの問題に対する局所探索アルゴリズムは現実的な時間内で高品質な解を生み出すことができるため、系統学の研究に対して重要な貢献を与える。

閉路完全距離の、それに関連する集合族による特徴付けは、重複クラスタリング及び系統樹推定問題の分野において理論的な基礎を与える。また、任意の相違写像を k -連結完全距離によって近似するための効率的なアルゴリズムは、NP困難な最適化問題に対する近似アルゴリズムの可能性を示唆する。

研究成果の概要（英文）：Given a dissimilarity mapping, to find an ultrametric which does not exceed and best approximates the dissimilarity matrix with respect to the L_p norm (the L_p -minimum increment ultrametric problem) is an important task in Phylogenetics. If p is infinite, then the problem can be solved by an efficient algorithm but if p is finite this problem is NP-hard. In this study, we give a local search algorithm for the L_p -minimum increment ultrametric problem for $p=1,2$. Cycle-complete distances are a generalization of ultrametrics. In this study, we give a characterization of cycle-complete distances in terms of the associated set families. Moreover, we introduce a generalization of cycle-complete distances called k -connected complete distances and give an algorithms for finding approximations of a given dissimilarity mapping by a k -connected complete distance with respect to the L_p norm.

研究分野：組合せ最適化

キーワード：組合せ最適化 超距離 木距離 アルゴリズム 系統樹推定問題

1. 研究開始当初の背景

系統学における重要な問題は、生物種間の相違を表わす行列（相違写像）が与えられたときに、これらの生物種の進化の歴史を忠実に表現する系統樹を推定することである。系統樹のモデルとして最も基本的なものは超距離木である。超距離木とは根から任意の葉への距離がすべて等しいような枝重み付き根付き木である。与えられた相違写像から系統樹（超距離木）を推定する方法の一つとして最適化によるアプローチがある。M を対象とする生物種に対する相違写像とする。系統樹推定問題に対する最適化によるアプローチでは、 $\|M - D_T\|_p$ を最小化するような超距離木 T を求める。ここで、 D_T は T の葉間距離を表わす行列であり、 $\|\cdot\|_p$ は L_p -ノルムである。この問題は最良近似超距離木問題と呼ばれ、 $p = \infty$ のときは多項式時間アルゴリズムが存在するが $p < \infty$ のときは NP-困難である [5]。また、入力として与えられた生物種間の相違は実際は真の相違の下限であるという場合がしばしばある。このような状況においては最良近似超距離木問題に対して $M \geq D_T$ という制約を追加したものを考える方がより適切である。この問題は最小増加超距離木問題と呼ばれ、最良近似超距離木問題の場合と同様に $p = \infty$ のときは多項式時間アルゴリズムが存在するが $p < \infty$ のときは NP-困難である [5]。 $p < \infty$ の場合の最良近似超距離木問題や最小増加超距離木問題に対して分枝限定法や近似アルゴリズムが提案されている。しかしながら、分枝限定法の適用可能性は入力行列の次数が 30 以下であるような問題のみに限定されている。また、近似アルゴリズムの精度は実用的に満足いくものではない。

超距離木に対応して超距離と呼ばれる相違写像の概念がある。近年になって超距離の概念は様々な形で一般化が成されているが、これらの一般化の中で特に本研究で着目するものが閉路完全距離 [6] である。閉路完全距離は、超距離の一般化であると同時に準超距離 (quasi-ultrametric) [2] と呼ばれる相違写像の部分クラスである。超距離がその木表現である超距離木に 1 対 1 に対応するのに対して、準超距離は指標付き準階層構造 (indexed quasi-hierarchy) と 1 対 1 に対応する。指標付き準階層構造は、重複クラスタリング (overlapping clustering) の一つのクラスとして見る事ができる。また、系統樹推定問題の文脈においても準階層構造によって生物種の進化をモデル化する研究が注目を集めている [3]。しかしながら、指標付き準階層構造は非常に一般的な概念であるために、超距離の持つような豊かな理論というものは存在していなかった。

2. 研究の目的

本研究は、超距離（木）の一般化でありながら超距離（木）の持つ「良い」性質を保存している準超距離（指標付き準階層構造）の部分クラスについて、その構造的性質を明らかにするとともに関連する最適化問題に対する効率的なアルゴリズムを与えることを目的とする。これによって重複クラスタリング及び新しい進化モデルへの応用に対して理論的基盤を提供することが可能になる。具体的には以下の 3 つの研究課題について研究を行なう。(1) 最良近似超距離木問題に対する局所探索アルゴリズムの開発、(2) 閉路完全距離のグラフ表現と閉路完全距離による相違写像の近似、及び、(3) k-連結完全距離による相違写像の近似理論の確立、である。以下では、これらの研究課題について説明する。

2.1 最良近似超距離木問題に対する局所探索アルゴリズムの開発

数理的最適化に対する一般的な手法の一つとして局所探索アルゴリズムがあるが、筆者の知る限りにおいて最良近似超距離木問題あるいは最小増加超距離木問題に対する局所探索アルゴリズムは存在しない。本研究ではこれらの問題に対する局所探索アルゴリズムを開発する。

2.2 閉路完全距離のグラフ表現と閉路完全距離による相違写像の近似

閉路完全距離は指標付き準階層構造として超距離木を一般化したグラフ的概念によって表現可能である。これを閉路完全距離グラフと呼ぼう。近年系統樹推定問題の文脈において、超距離（木）よりも現実の進化過程を記述する能力が高いモデルの研究が活発に行われている [3]、筆者は閉路完全距離（グラフ）が超距離（木）を一般化した進化のモデルの有望な候補であると考えている。この一般的な進化のモデルにおいて最も基本的な問題は最良近似閉路完全距離問題である。すなわち、与えられた相違行列を最も良く近似する閉路完全距離を見出す問題である。L_p-ノルムに関して最良近似閉路完全距離問題については、Ando et al. [1] によって定義された subdominant 閉路完全距離を用いることによって最適解を得ることが可能であるが、この問題は有限の p に対する L_p-ノルムに関しては未解決の問題である。本研究では有限の p に対する最良近似閉路完全距離問題に対するアルゴリズムを開発することを目標とする。

2.3 k-連結完全距離による相違写像の近似理論の確立

前述したように閉路完全距離は超距離の一般化である．この一般化がどのように行われたかということに注意深く観察することによって閉路完全距離のさらなる一般化が自然に導かれる．これを k -連結完全距離と呼ぼう． k -連結完全距離とそれに対応するグラフ構造は，閉路完全距離（グラフ）よりもさらに高い記述能力を持つ進化のモデルとして有用であると考えられる．さらに， k -連結完全距離は重複クラスタリングにおける指標付き準階層構造の良い性質を持った部分クラスを構成していると予想している．任意の相違写像に対して subdominant 超距離（閉路完全距離）が存在して，それが与えられた相違写像を L_1 -ノルムの意味で最良近似することが知られている．本研究では，この性質は k -連結完全距離においてもそのまま成り立つということを示し，さらに，任意の k に対して subdominant k -連結完全距離を見出す効率的なアルゴリズムを開発する．

3．研究の方法

3.1 最良近似超距離問題に対する局所探索アルゴリズムの開発

最小増加超距離木問題は，重みが与えられていない2分木を全探索することによって解くことが可能であることを示すことができる．したがって，2分木に対して適切な近傍概念を定義することによって局所探索アルゴリズムの設計が可能となる．系統樹推定問題の一つである最大節約問題（maximum parsimony problem）に対して SPR 近傍と呼ばれる2分木の近傍概念に基づく局所探索アルゴリズム [8] が知られている．また，与えられた相違行列を最も良く近似する（超距離木とは限らない）枝重み付き木を求める問題（最良近似木問題）の $p = 2$ の場合に対しては，NNI 近傍に基づく局所探索アルゴリズム [7] が提案されている．本研究では，これらの2分木の近傍概念を用いて最良近似超距離木問題と最小増加超距離木問題に対する局所探索アルゴリズムを開発する．さらにその有用性を数値実験によって検証する．

3.2 閉路完全距離のグラフ表現と閉路完全距離による相違写像の近似

本研究ではまず，与えられた閉路完全距離に対してそれを表現する閉路完全距離グラフを構成する多項式時間アルゴリズムを開発する．このアルゴリズムは，Ando et al. [1] において与えた subdominant 閉路完全距離を求めるアルゴリズムを変形することによって構成が可能であると予想している．このアルゴリズムの完成後，計算機で自動生成された閉路完全距離グラフを観察することによって閉路完全距離グラフの特徴付けを与える．次に，閉路完全距離による相違関数の近似アルゴリズムとして局所探索アルゴリズムを提案する．このアルゴリズムでは，最良近似超距離木問題に対する局所探索アルゴリズムで用いられたアイデアが用いられる．すなわち，この局所探索アルゴリズムで用いられる近傍として閉路完全距離グラフに対する局所的変形操作によって定義されるものを用いる．さらに，開発したアルゴリズムの性能を評価するために計算機実験を行う．

3.3 k -連結完全距離による相違写像の近似理論の確立

L_∞ -ノルムに関する最良近似超距離問題は subdominant 超距離 [4] を用いることによって多項式時間で解くことが可能である．Chepoi et al. [4] は，subdominant の概念を超距離を含むより広い概念へと拡張している．筆者は，Chepoi et al. の理論に基づいて L_∞ -ノルムに関する最良近似 k -連結完全距離問題も，これと同様に subdominant k -連結完全距離を用いることで解くことができることを示す．さらに，subdominant k -連結完全距離を求める効率的なアルゴリズムを開発し，その性能を計算機実験によって評価する．

4．研究成果

4.1 閉路完全距離に関連する2-階層構造による特徴付け

超距離は指標付き階層構造との間に1対1対応が存在し，超距離の一般化である2-超距離は指標付き2-階層構造との間に1対1対応が存在する．閉路完全距離は超距離の一般化であり，かつ，2-超距離の部分クラスである．したがって，閉路完全距離は指標付き2-階層構造の部分クラスに対応する．本研究では，この指標付き2-階層構造の部分クラスを，指標付き「無閉路」2-階層構造として特徴付けた．この特徴付けによって閉路完全距離の特徴付けが導かれる．さらに，任意の閉路完全距離が与えられたときに，対応する無閉路指標付き2-階層構造を出力する $O(n^2 \log n)$ 時間アルゴリズムを与えた．ここで， n は与えられた閉路完全距離が定義される台集合のサイズである．

4.2 閉路完全距離に関連する2-階層構造に対する凝集的クラスタリング

階層クラスタリングとは，データの集合をいくつかの集合に分割するクラスタリング手法の一種である．この手法は単元集合だけからなる状態から出発し，クラスタ間距離が最も近い2つのクラスタを併合する．この併合をすべての要素が1つのクラスタに含まれるまで繰り返し，クラスタの階層を生成する．階層クラスタリングにはクラスタ間距離の定義の違いにより，単純連結法や完全連結法などがある．階層クラスタリングによって生成されるクラスタの階層は指

標付き階層構造であり、指標付き階層構造には超距離と呼ばれる相違写像のクラスが一对一に対応する。閉路完全距離には指標付き無閉路 2-階層構造が一对一に対応する。指標付き階層構造の各階層は与えられた台集合の分割であるのに対して、指標付き無閉路 2-階層構造の各階層にある異なる 2 つの集合は交わりを持つことが許される。本研究では階層クラスタリングの代表的なアルゴリズムである単純連結法と完全連結法を拡張して、指標付き無閉路 2-階層構造を出力する階層重複クラスタリングアルゴリズムを開発した。

4.3 最小増加超距離問題に対する局所探索アルゴリズムの開発

$p=1$ 及び 2 に対する L_p 最小増加超距離木問題に対する部分木交換操作と呼ばれる 2 分木の变形操作に基づく局所探索アルゴリズムの効率的な実装方法を開発した。さらに数値実験によって、このアルゴリズムが既存の実装方法よりも理論的にも実際的にも効率的であるということを示した。

4.4 Subdominant k -連結完全距離の特徴付けとそれに対するアルゴリズム

閉路完全距離の自然な一般化として、任意の正整数 k に対して k -連結完全距離の概念を導入し、任意の相違写像 d に対してその subdominant k -連結完全距離が存在することを証明した。さらに、subdominant k -連結完全距離の特徴付けを与え、その特徴付けに基づいて、任意の相違写像に対してその subdominant k -連結完全距離を求める $O(kn^4)$ 時間アルゴリズムを与えた。また、任意の相違写像に対する連結完全度の概念を導入し、連結完全度を求める $O(n^5)$ 時間アルゴリズムを与えた。

4.5 部分木距離を表現する木を求めるアルゴリズムの開発

超距離木を緩和した系統樹の概念として枝重み付き木があり、これに対応する相違写像が木距離である。この一般化に関連する最適化問題として、与えられた相違行列を木距離で近似する問題が超距離木最適化問題と同様に定義される。ここで最も基本的な問題は、任意の相違写像が与えられたときにそれが木距離かどうかを判定し、もしそうであればそれを表現する枝重み付き木を求める問題である。この問題は古くから多くの研究成果がある。部分木距離とは木距離を一般化した概念であり、重み付き木とその部分木の集合によって定義される。本研究では、与えられた任意の相違写像が部分木距離であるかを判定し、もしそうであればそれを定義する重み付き木とその部分木の集合を求める最適な時間計算量を持つアルゴリズムを開発した。

< 引用文献 >

- K. Ando, R. Inagaki and K. Shoji: Efficient algorithms for subdominant cycle-complete cost functions and cycle-complete solutions. *Discrete Applied Mathematics*, Vol. 225, 2017, 1-10.
- H.-J. Bandelt and A.W.M. Dress: An order theoretic framework for overlapping clustering, *Discrete Mathematics*. Vol. 136, 1994, 21-37.
- U. Brandes and S. Cornelsen: Phylogenetic graph models beyond trees. *Discrete Applied Mathematics*, Vol. 157, 2009, 2361-2369.
- V. Chepoi and B. Fichet: L -approximation via subdominants. *Journal of Mathematical Psychology*, Vol. 44, 2000, 600-616.
- Farach, K. Kannan and T. Warnow: A robust model for finding optimal evolutionary trees. *Algorithmica*, Vol. 13, 1995, 155-179.
- C. Trudeau: A new stable and more responsive cost sharing solution for minimum spanning tree problems. *Games and Economic Behavior*, Vol. 75, 2012, 402-412.
- R. Desper and O. Gascuel: Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology*, Vol. 9, 2002, 687-705.
- A. Goeffon et al.: Progressive tree neighborhood applied to the maximum parsimony problem. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 5, 2008, 136-145.

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Takanori Maehara and Kazutoshi Ando	4. 巻 E105-A
2. 論文標題 Optimal algorithm for finding representation of subtree distance	5. 発行年 2022年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 1203 ~ 1210
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transfun.2021DMP0014	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kazutoshi Ando Noriyoshi Sukegawa and Shota Takagi	4. 巻 65
2. 論文標題 Strong Condorcet criterion for the linear ordering problem	5. 発行年 2022年
3. 雑誌名 Journal of the Operations Research Society of Japan	6. 最初と最後の頁 67 ~ 75
掲載論文のDOI (デジタルオブジェクト識別子) 10.15807/jorsj.65.67	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Ando Kazutoshi, Fujishige Satoru	4. 巻 36
2. 論文標題 Signed ring families and signed posets	5. 発行年 2021年
3. 雑誌名 Optimization Methods and Software	6. 最初と最後の頁 262 ~ 278
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10556788.2020.1740219	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Ando Kazutoshi, Takase Koichi	4. 巻 63
2. 論文標題 MONTE CARLO ALGORITHM FOR CALCULATING THE SHAPLEY VALUES OF MINIMUM COST SPANNING TREE GAMES	5. 発行年 2020年
3. 雑誌名 Journal of the Operations Research Society of Japan	6. 最初と最後の頁 31 ~ 40
掲載論文のDOI (デジタルオブジェクト識別子) 10.15807/jorsj.63.31	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ando Kazutoshi, Shoji Kazuya	4. 巻 11346
2. 論文標題 Characterizing Cycle-Complete Dissimilarities in Terms of Associated Indexed 2-Hierarchies	5. 発行年 2018年
3. 雑誌名 Lecture Notes in Computer Science	6. 最初と最後の頁 640 ~ 650
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-04651-4_43	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件 (うち招待講演 0件 / うち国際学会 1件)

1. 発表者名 安藤和敏, 水越雅紀
2. 発表標題 l2最小増加超距離木問題に対するk制限部分木交換近傍に基づく局所探索アルゴリズム
3. 学会等名 日本オペレーションズ・リサーチ学会2022年春季研究発表会
4. 発表年 2022年

1. 発表者名 安藤和敏, 杉本達哉, 鮎川矩義
2. 発表標題 分解定理に基づく線形順序付け問題の発見的解法
3. 学会等名 日本オペレーションズ・リサーチ学会2021年秋季研究発表会
4. 発表年 2021年

1. 発表者名 安藤和敏, 水越雅紀
2. 発表標題 最小増加超距離木問題に対する最小増加超距離木問題に対するk制限部分木交換近傍に基づく局所探索アルゴリズム
3. 学会等名 日本オペレーションズ・リサーチ学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 水越雅紀, 安藤和敏
2. 発表標題 最小増加超距離木問題に対する高速NNI アルゴリズム
3. 学会等名 京都大学数理解析研究所研究集会「数理最適化の理論・アルゴリズム・応用」
4. 発表年 2021年

1. 発表者名 安藤和敏, 松田泰比都
2. 発表標題 閉路完全距離に対応する凝集的階層重複クラスタリング
3. 学会等名 日本オペレーションズ・リサーチ学会2020年春季研究発表会
4. 発表年 2019年

1. 発表者名 安藤和敏, 鮎川矩義, 高木祥多
2. 発表標題 強Condorcet規準とKemeny順序
3. 学会等名 日本オペレーションズ・リサーチ学会2019年秋季研究発表会
4. 発表年 2020年

1. 発表者名 安藤和敏, 正治和也
2. 発表標題 Subdominant k 連結完全距離の特徴付けとそれに対するアルゴリズム
3. 学会等名 日本オペレーションズ・リサーチ学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 K. Ando, K. Shoji
2. 発表標題 Characterizing Cycle-Complete Dissimilarities in Terms of Associated Indexed 2-Hierarchies
3. 学会等名 International Conference on Combinatorial Optimization and Applications, COCOA 2018 (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	正治 和也 (Shoji Kazuya)		
研究協力者	高瀬 光一 (Takase Koichi)		
研究協力者	松田 泰比都 (Matsuda Yasuhiro)		
研究協力者	水越 雅紀 (Mizukoshi Masaki)		
研究協力者	鮎川 矩義 (Sukegawa Noriyoshi) (20757710)	筑波大学・システム情報系・助教 (12102)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	高木 祥多 (Takagi Shota)		
研究協力者	藤重 悟 (Fujishige Satoru) (10092321)	京都大学・数理解析研究所・名誉教授 (14301)	
研究協力者	杉本 達哉 (Sugimoto Tatsuya)		
研究協力者	前原 貴憲 (Maehara Takanori)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関